# DRIVING CYCLES: A NEW CYCLE-BUILDING METHOD THAT BETTER REPRESENTS REAL-WORLD EMISSIONS

Task Order No. 66

September 18, 2008

By

**Zhen Dai**[*]
**Deb Niemeier, PhD, PE**
**Douglas Eisinger, PhD**

## Abstract

**Background**: On-road vehicle emissions models couple emission rates with travel activities. Emission rates are derived from data collected using dynamometers. Dynamometer tests measure emissions while a vehicle follows a drive cycle with various speed and modal (acceleration, cruise, deceleration, idle) events. Existing cycle construction methods focus on representing real-world activity, without also directly weighing the relationship between activity and emissions. In addition, cycles have traditionally focused on representing complete trips, rather than activity on specific roadway links or road types.

**Methods**: This paper presents a new cycle construction methodology that offers three advantages over traditional approaches. First, it creates cycles that represent statistically-defined speed classes; this groups activities into bins more closely associated with emissions. Second, it defines modal events using speed and event-duration criteria that improve representation of real-world activity. Third, it improves upon traditional approaches that create cycles primarily to match real-world Speed-Acceleration Frequency Distributions (SAFDs). The new method requires that cycles match real-world SAFDs *and* distributions of modal events. This work also introduces a new test statistic to assess cycle performance: the Composite Performance Measure (CPM). The CPM measures how well a cycle matches the SAFD and modal distributions of real-world data. We illustrate these new methods by creating arterial roadway cycles.

**Results**: Arterial driving can be represented by four activity clusters: < 12.5 mph, 12.5 - 32.5 mph, 32.5 - 47.5 mph, and > 47.5 mph. Cycles representing the two mid-range speed activity clusters performed equally well regardless of whether they were created with the new techniques or with a traditional approach. However, cycles created for the lowest and highest speed activity clusters performed better at representing real-world data when they were constructed to match SAFD *and* mode distributions, rather than SAFD metrics alone. The percentage difference between observed and cycle-constructed mean accelerations was a factor of two to three smaller for the low- and high-speed cycles created using SAFD and mode distribution metrics.

---

[*]Formerly with the U.C. Davis-Caltrans Air Quality Project; currently with the California Air Resources Board.

# *About* The U.C. Davis-Caltrans Air Quality Project

**http://AQP.engr.ucdavis.edu/**

Department of Civil & Environmental Engineering
University of California
One Shields Ave., Davis, CA 95616
(530) 752-0586

**Mission**:  The Air Quality Project (AQP) seeks to advance understanding of transportation related air quality problems, develop advanced modeling and analysis capability within the transportation and air quality planning community, and foster collaboration among agencies to improve mobility and achieve air quality goals.

**History**: Since the 1990s, the U.S. Federal Highway Administration and Caltrans have funded the AQP to provide transportation-related air quality support.  Caltrans and AQP researchers identify and resolve issues that could slow clean air progress and transportation improvements.

**Accessibility**: AQP written materials and software tools are distributed through our website, peer-reviewed publications, conference presentations, training classes, formal reports and technical memoranda, and periodic newsletters.

**Research**:  AQP investigations focus on project-level, regional-scale, and national-level assessments.  Tools and publication topics cover pollutant-specific problems such as those involving particulate matter, carbon monoxide, carbon dioxide, ozone and air toxics; activity data collection and assessment for on- and off-road vehicles and equipment; mitigation options such as transportation control measures; policy analyses addressing transportation conformity and state implementation plan development; litigation support; and goods movement assessments.

## Project Management

Principal Investigator and Director:  Deb Niemeier, PhD, PE
Program Manager:  Douglas Eisinger, PhD

Caltrans Project Manager:  Mike Brady, Senior Environmental Planner
Air Quality and Conformity Coordination
Division of Transportation Planning, MS-32
California Department of Transportation
1120 N Street, Sacramento, CA 94274
(916) 653-0158

# ACKNOWLEDGMENTS

**TABLE OF CONTENTS**

# 1. Introduction

In the United States, models such as EMFAC (in California) and MOBILE (in the rest of the U.S.) estimate on-road vehicle emissions by coupling emission rates with corresponding activities. Mobile emission rates are generally developed from dynamometer tests using driving cycles such as Federal Test Procedure (FTP) cycles or the California Unified Cycle (Austin et al. 1995). Consequently, the representativeness of driving cycles is crucial for emission assessment.

A driving cycle is a speed-time profile designed to represent a real-world driving pattern. Cycle length ranges from one or two minutes to 30 minutes or more; duration is typically limited to reduce per-vehicle test costs. Thus, cycle development efforts are challenged to mimic real-world behavior using limited duration activities. During a dynamometer test, a vehicle follows a specific speed-time trace while emissions and energy consumption are recorded. Using dynamometer data, emission rates or speed correction factors are then developed for the particular driving activity under study and existing emissions models incorporate the findings to represent emissions under different driving conditions.

Most cycle construction methods, meaning those traditionally used to support development of the EMFAC and MOBILE emissions models, have focused on representing real-world driving activity, rather than real-world emissions. This paper presents a new cycle construction methodology designed to differentiate driving patterns based on their emissions, and to represent driving activity and associated emissions from a modal perspective.

# 2. A Review of Basic Methodologies of Cycle Construction

In general, cycle construction methods typically include the following steps: (1) collecting real world driving data, (2) segmenting the driving data, (3) constructing cycles, and (4) evaluating and selecting the final cycle (Andre 2004). Depending on the type of driving activity that is being using to construct the cycle, existing cycle construction methodologies for light-duty vehicles can be generalized into four types: micro-trip cycle construction; trip segment-based cycle construction, cycle construction based on pattern classification, and modal cycle construction.

## 2.1   Micro-trip based cycle construction

One approach to create driving cycles is to chain "micro-trips," which are defined as the driving activity between adjacent stops, including the leading period of idle (Austin et al. 1993). Examples include the Sydney cycle (Kent et al. 1978), the Melbourne Peak Cycle (Watson et al. 1982), the Unified Cycles (Austin et al. 1993), the Unified Correction Cycles (Gammariello and Long 1996), and the Hong Kong driving cycles (Hung et al. 2005).

With the micro-trip method, real world activity data, for example, the modal activity of numerous vehicles tracked for several minutes each while driving on a freeway or an arterial, is divided into micro-trips. A driving cycle is constructed by chaining representative micro-trips with the goal that the cycle closely matches the observed data. Micro-trips are usually selected using one of three methods: random selection; "best incremental," meaning incrementally searching for and adding a micro-trip with specific modal characteristics; or a hybrid of both approaches (Austin et al. 1993). Candidate cycles are assessed using parameters that typically include average speed, maximum speed, minimum speed, average acceleration, average deceleration and the Speed-Acceleration Frequency Distribution (SAFD).

The major limitation to the micro-trip based methods is that the micro-trip does not differentiate by various types of driving conditions such as roadway type or Level of Service (LOS). As a result, replicating driving activities under a particular driving condition is difficult (Andre 2004). For example, under smooth traffic conditions, a vehicle seldom stops and a single micro-trip may cover different road segments or different traffic conditions. Use of micro-trip based methods has been limited to developing cycles designed to represent a single type of trip or cycles designed to replicate region-wide driving conditions.

## 2.2   Segment-based cycle construction

A trip "segment" is obtained by partitioning vehicle speed-time profiles using changes in roadway type or LOS, in addition to stops (Carlson and Austin 1997). Therefore, overall vehicle driving activity can be stratified by roadway type or LOS, and cycles can be constructed to represent driving activity for specific roadway types and traffic conditions. One major application of the segment-based cycle construction method is U.S. Environmental Protection Agency (EPA) facility-specific speed correction cycles.

Like the micro-trip based method, trip segments are selected and concatenated using a hybrid of random and best-incremental logic. However, unlike a micro-trip, trip segments can start and end at any speed. Therefore, the chaining of segments requires

certain constraints on speed and acceleration between two connecting seconds respectively of previous and succeeding micro-tips. For instance, in constructing EPA's facility specific cycles, the differences in speeds between two connecting seconds was required to be within 0.5 mph, and the difference in acceleration within 0.5 mph/sec (Carlson and Austin 1997). The "best" cycle is then selected using two primary parameters: 1) the sum of difference in SAFD between the test cycle and the target population, and 2) the amount of operation occurring in high power mode (Carlson and Austin 1997).

The limitation of segment-based cycle construction is that data are stratified strictly from a transportation engineering perspective and, consequently, are less related to emissions.

*2.3    Cycle construction with pattern classification*

This type of method has been widely applied in constructing European driving cycles (Andre 1996; Andre et al. 1995; Andre and Rapone 2008). Under this approach, "kinematic sequences" (similar to micro-trips) are classified into heterogeneous classes using statistical methods. Also, the approach uses succession probabilities to estimate and consider the likelihood that one class of activity precedes or follows a different activity class. Driving cycles are constructed by re-connecting kinematic sequences randomly selected from each of the activity classes in accordance with the probability and chronology of kinematic sequences (Andre et al. 1995).

In constructing the European Urban Cycles (Andre et al. 1995), kinematic sequences are described by 20 variables including duration, idle time and distance throughout the sequence; the means, maximums, standard deviations, 20% and 80% percentiles of driving modes; and other measures such as instantaneous speed and acceleration, and the distance between two accelerations. Principal Component Analysis (PCA) was applied to these 20 variables, followed by cluster analysis on the kinematic sequences. Four distinctive classes were identified, respectively representing congested and free-flow urban traffic, extra-urban and motorway driving conditions.

Next, a trip was viewed as a series of kinematic sequences. All observed trips were classified based on the frequency of sequences in each kinematic sequence class and the number of transitions between two classes of kinematic sequences. Three major types of trips were identified, namely urban trips, road trips and motorway trips.

The method was further improved in constructing the European driving cycles under the ARTEMIS project (Andre 2004). To avoid bias due to the varying duration of sequences between two stops, speed-time traces were segmented into sequences of homogeneous

size rather than partitioned by stops.  Correspondence Analysis using chi-squared distance and cluster analysis were applied to classify the sequences into 12 driving conditions.  Urban, rural road and motorway cycles were developed based on the observed composition of driving conditions.

The limitations of this method are that: 1) the basic units of cycle construction, including micro-trip and uniform sequences, are not directly related to emissions, and potentially not the best units to be used in defining emission-related driving activity, and 2) the classification of sequences is based on the chi-square distance of speed-acceleration joint distribution.  Although such classification differentiated the kinematic driving activity, it does not necessarily differentiate emissions associated with these activities.

## 2.4    Modal cycle construction

A study by Lin and Niemeier (2002) applied a mode-based cycle construction method, where real world driving is viewed as a sequence of acceleration, deceleration, cruise, or idle modes.  Studies show that running emissions are related to vehicle modal operation in addition to average speed (An et al. 1998; Barth et al. 1996; Fomunung et al. 1999).  Therefore, for emission estimation purposes, it is logical to analyze and replicate driving activities from a modal perspective.  Assuming that the likelihood of a particular modal event (e.g., acceleration, cruise, or deceleration) occurring depends only on the mode of the previous modal event, driving activities can be modeled as a Markov Chain.

Modal cycle construction comprises four basic steps.  First, the real-world driving data is partitioned into snippets of various durations based on acceleration using a maximum likelihood estimation (MLE) clustering method (Lin and Niemeier 2002).  Second, the snippets are classified into different modal bins, again using the MLE clustering method.  This time, the clustering variables include average, minimum, and maximum speeds and acceleration rates.  The third step creates a transition matrix that contains the succession probabilities between different modes.  Finally the cycle is constructed as a Markov chain; to add one additional snippet, the next modal bin is predicted based on the modal nature of the current snippet and the transition matrix.  One snippet is selected from the predicted bin without replacement.  The snippet selection requires that the selected snippet best improves the match to the observed SAFD, and that the start speed of the snippets matches the end speed of the previous modal snippets with an acceptable difference (within  0.2 km/h in Lin's study).  Snippet selection is repeated until the desired cycle length is achieved.  The final driving cycles are selected using a composite assessment measure, which integrates parameters such as differences in average speed, differences in minimum and maximum speeds, and percentage of idling operation.

4

The shortcomings of the modal approach include: 1) the criterion used to connect the snippets is arbitrary, and 2) cycles are constructed for a specific facility and LOS, and the number of cycles needed to represent the emission related driving activity are not well studied (Lin and Niemeier 2003). The cycle construction method introduced here builds upon the work done by Lin and Niemeier.

## 3. New Modal Driving Cycle Construction

### 3.1 Basic principles

We improve the ability of driving cycles to represent real world emissions by including two major steps in cycle construction process: 1) a dynamic statistical analysis of driving patterns and 2) cycle synthesis for specific driving patterns.

The goal of the first step, statistical analysis, is to differentiate real world driving activities in terms of their emission-producing capability, and in doing so to determine the total number and individual characteristics of desired driving cycles. To develop an improved methodology, we used driving data obtained by chase-vehicles deployed on California arterial roadways. We applied Principal Component Analysis (PCA) and cluster analysis to characterize emissions-related driving activity. The results of these analyses then served as a benchmark against which we later assessed the quality of the driving cycles we created using our new cycle development methodology.

The objective of the second step, cycle synthesis, is to develop a modal cycle construction method which reflects the relationship between emissions and vehicle operation modes, and retains the chronology of real-world modal events. We developed our methodology by expanding Lin and Niemeier's (2002) Markov Chain snippet-concatenating concept. Improvements were made to data segmentation and the boundary condition in snippet concatenation.

The new method partitions continuous driving activity into different modal bins based on both the intensity and duration of speed and acceleration. With the new data segmentation method, we successfully partitioned real-world vehicle speed data (i.e., second-by-second traces), into modal snippets of acceleration, deceleration, cruise and idle. These modal snippets constitute basic working units in our cycle construction.

*3.2 Data source*

For this study, we used target vehicle data collected by the Caltrans/ARB Modeling Program (CAMP) using a chase car method. The dataset contains second-by-second light-duty vehicle driving data for different types of roadways under different congestion levels in four differentially urbanized areas: the Metropolitan Bay Area, Sacramento, Stanislaus and the South Coast area. The driving routes were designed to be representative not only from the trip perspective (i.e., origin-destination pairs, trip length), but also from the facility perspective (e.g., proportion of VMT on freeway or arterials) (Eisinger et al. 2006).

To construct new arterial cycles, we used the portion of CAMP data collected on arterials. Given the difference in chase vehicle and target vehicle driving profiles (Morey et al. 2000), we used only target vehicle data to construct arterial cycles. In total, we have over 330,000 seconds of observed arterial driving data.[1]

*3.3 Characterizing driving patterns*

<u>Overview</u>

The first step of our cycle construction requires that driving activity be classified with respect to emissions producing capability. To do this, we searched for variables that best captured the emissions associated with driving activities. We performed PCA on these variables to generate "distance measures" which, through cluster analysis, were used to identify unique groups in the observed data. A "distance measure" refers to the degree to

---

[1] The raw arterial driving data contained a few seconds of abnormally large accelerations and decelerations. The largest acceleration recorded was 45 mph/s and largest deceleration 48 mph/s. Because a vehicle's acceleration performance is constrained by engine power and vehicle weight, such modal operations are unlikely to be real. Existing studies have observed accelerations within a range of +/- 15 mph/s (e.g., (Goodwin and Ross 1996; Watson et al. 1982). Moreover, were such extreme values to be included in a driving cycle, it would be impossible for a test vehicle to follow the trace containing the aberrant data points. In examining the data, we found that all of the abnormally large values were recorded by a single chase vehicle during a specific time period in the Bay Area. We believe that the observations reflect measurement error associated with the specific equipment on that vehicle. To filter out the equipment error outliers, we used the maximum deceleration and acceleration recorded in any other region (+/- 22 mph/s) as cut-points. Observations with greater deceleration or acceleration were treated as outliers and removed from the analysis. These excluded data constitute less than 0.1% of the total arterial target vehicle observations.

which groups of data differ from one another – as distance measures increase, data groups are more distinct from each other.

Existing studies have suggested that the operating mode distribution is a robust measure for characterizing driving activities producing emissions (EPA 2005). To test the similarity (or dissimilarity) between regions, facilities, traffic speeds, or chase vehicles, we divided the complete second-by-second dataset by region, facility type, average speed and chase vehicle designation. The chase vehicle data did not record roadway link average speed. Average speed is a primary variable important in differentiating emissions, but is not the same as the instantaneous speed which is recorded by the chase vehicle. As an alternative, we calculated the average speed of both chase vehicle and target vehicles as a surrogate for link average speed. The average speeds were classified into bins starting from 5 mph, 10mph, …, up to 50 mph (Table 1).

**Table 1 Average speed bin definition**

| Average Speed Bin (mph) | (Surrogate) Average Speed (mph) |
|:-----------------------:|:-------------------------------:|
| 5                       | 0 ~ 7.5                         |
| 10                      | 7.5 ~ 12.5                      |
| 15                      | 12.5 ~ 17.5                     |
| …                       | ...                             |
| 50                      | > 47.5                          |

From each subset of data, we generated the operating mode distribution, that is, the distribution of time by operating mode. Operating modes are defined by instantaneous speed, Vehicle Specific Power (VSP) and acceleration (refer to Table 2). We computed VSP for each second of vehicle arterial driving data, using parameters provided by EPA's MOVES model (EPA 2005):

$$VSP = (A * v + B * v^2 + C * v^3 + M * v * a) / M$$

Where,

$$
\begin{aligned}
VSP &= \text{vehicle specific power in kW/Metric Ton} \\
v &= \text{the instantaneous speed in meters/second (m/s)} \\
a &= \text{the instantaneous acceleration in meters/second}^2 \\
A &= \text{the rolling resistance term in kW·s/m} \\
B &= \text{the friction term in kW· s}^2 / \text{m} \\
C &= \text{the aerodynamic drag term in kW·s}^3 / \text{m} \\
M &= \text{vehicle mass in metric tons (1000 kg)}
\end{aligned}
$$

Each second of observed data was assigned to a specific operating mode bin (Table 2).

**Table 2 Operating mode bin definitions for MOVES 2004**

Source: (EPA 2004) Table 9-4.

| Braking (Bin 0) | | | |
|---|---|---|---|
| Idle (Bin 1) | | | |
| VSP | Speed | | |
| | 0-25mph | 25-50mph | >50mph |
| <0 KW/tonne | Bin 11 | Bin 21 | Bin 33 |
| 0 to 3 | Bin 12 | Bin 22 | |
| 3 to 6 | Bin 13 | Bin 23 | |
| 6 to 9 | Bin 14 | Bin 24 | Bin 35 |
| 9 to 12 | Bin 15 | Bin 25 | |
| 12 and greater | Bin 16 | Bin 26 | Bin 36 |

_PCA_

The operating mode distribution shown in Table 2 is a 17-dimension variable. We applied PCA on the operating mode distributions to reduce the data dimensions and to obtain distance measures between different operating mode distributions.  The results of the PCA showed that the first two principal components explained more than 85% of the variance in the data (Figure 1).  The first principal component is positively loaded from high-speed operating mode bins while negatively loaded from low-speed operating mode bins (Figure 2a). *This implies that the first principal component is associated with the mean speed*.  The second principal component is positively loaded from both high-speed and low-speed operating mode bins while negatively loaded from the medium-speed operating mode bins (Figure 2b), suggesting that *the second principal component is related with the degree that speeds deviate from the mean speed* (in other words, the low- and high-speed "tails" surrounding mean speeds).  Figure 3 confirms that both principal components are strongly related with average speeds.  From principal component plots (see Figures 3 and 4) for the first two principal components by location, we saw no distinctive pattern associated with a specific region.  In other words, the two main principal components do not seem to be strongly representative of any regional effect. This suggests that the regional effect, if any, may not be highly influential relative to the overall variance of driving patterns.  Based on a scree plot, we elected to use only the first two components in the next step, cluster analysis.
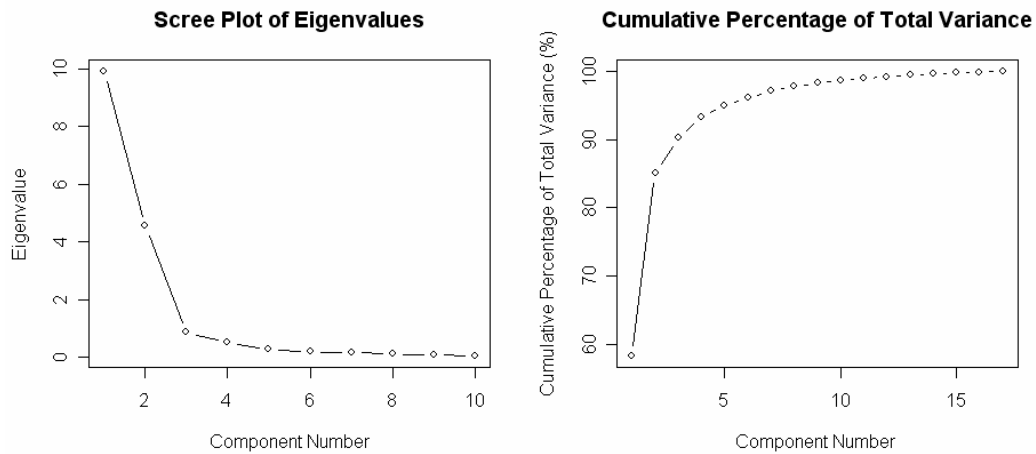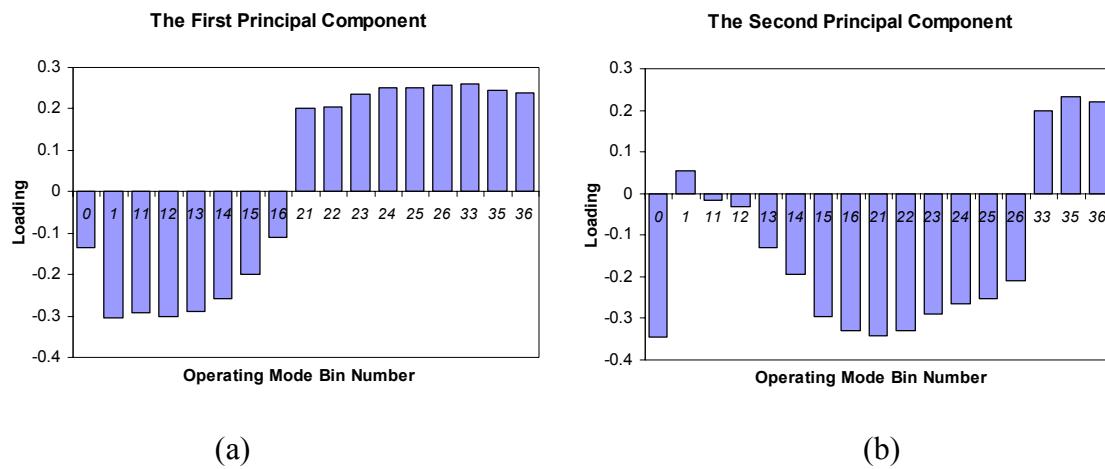
**Figure 1 Principal Component Analysis on the Operating Mode Distribution**



(a)                                                          (b)
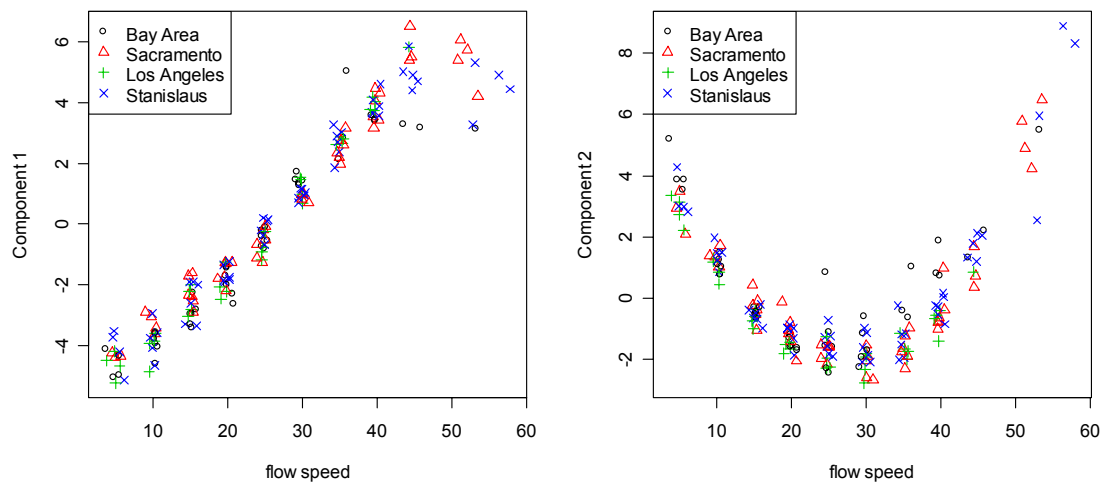
**Figure 2 Principal Component Loadings**

9

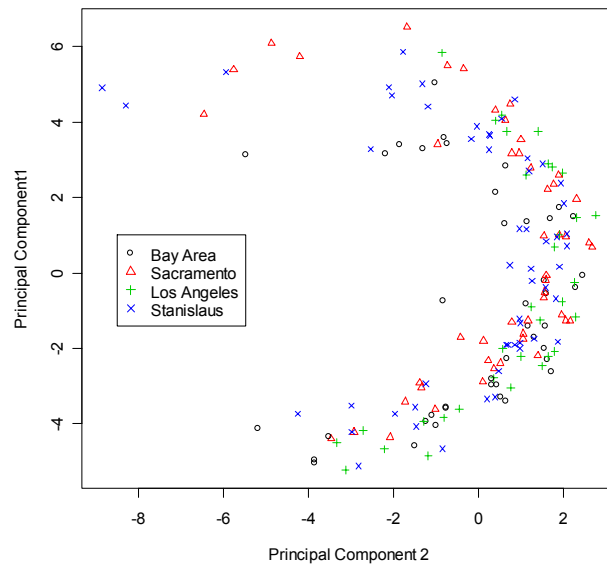**Figure 3 Plot of Principal Component Against Average Speeds**



**Figure 4 Principal Component Plot**

_Cluster analysis_

The principal component scores for the two principal components (one related to mean speed, one related to the deviation in speed) are used to calculate a "distance measure" between driving patterns (in other words, as a measure of how distinct different driving

patterns are from each other, considering variables that affect emissions). We applied cluster analysis to determine the number of homogeneous groups of driving patterns. We visually compared the results from a number of different clustering methods, including the average linkage method, single linkage method, Ward's Minimum Variance, Centroid method, and the mode-based method in SAS as well as the two-step clustering method in SPSS[2]. Based on visual examination, the two-step SPSS clustering method, the Centroid method and Average Linkage clustering method provided similar as well as the most intuitively reasonable results (see Figure 5 as an illustration of clustering results).

Finally, we used the two-step clustering method in SPSS to determine the number of clusters. A major advantage to this method is that the number of clusters can be decided based on the Bayesian Information Criteria (BIC) and the increase in distance measure (SPSS 2001):

$$BIC_M = -2L_M(x,\widehat{\theta}) + m\log(n)$$

where $L_M(x,\widehat{\theta})$ is the maximum log-likelihood for model M, $m$ is the number of parameters to be estimated and n is the number of observations. The second term, $m\log(n)$, adds a penalty to increasingly complex models.

The two-step clustering method produced four clusters. The relationship between average speed bin and the cluster membership is shown in Table 3. It is clear that each cluster corresponds well with a specific range of average speeds. For instance, the first cluster represents driving patterns of very slow moving vehicles (under 12.5 mph). The second cluster consists mostly of observations with average speeds between 12.5 mph to 32.5 mph. The third cluster consists of observations with higher average speeds between 32.5 mph to 47.5 mph. And the last cluster includes observations with the highest arterial average speeds (greater than 47.5 mph). Next, we construct cycles using subsets of driving data for each cluster.

---

[2] SAS and SPSS are commercially available software packages for completing statistical analyses.
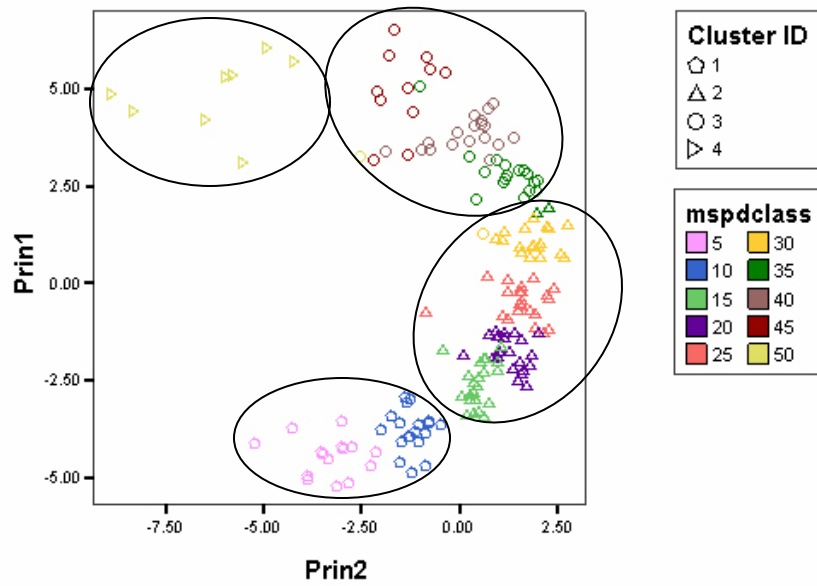
**Figure 5 Example Driving Pattern Clustering Result Based on the Centroid Method**

Note: the four clusters shown in Figure 5 are associated with specific average speed classes (see Table 3).

**Table 3 Cross-table of average speed class and cluster membership**

| Average Speed Class | Range of Average Speed (mph) | Number of observations in each cluster[*] | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| 5 | 0 ~ 7.5 | **15** | 0 | 0 | 0 |
| 10 | 7.5 ~ 12.5 | **19** | 0 | 0 | 0 |
| 15 | 12.5 ~ 17.5 | 0 | **22** | 0 | 0 |
| 20 | 17.5 ~ 22.5 | 0 | **22** | 0 | 0 |
| 25 | 22.5 ~ 27.5 | 0 | **22** | 0 | 0 |
| 30 | 27.5 ~ 32.5 | 0 | **19** | 1 | 0 |
| 35 | 32.5 ~ 37.5 | 0 | 2 | **17** | 0 |
| 40 | 37.5 ~ 42.5 | 0 | 0 | **19** | 0 |
| 45 | 42.5 ~ 47.5 | 0 | 0 | **11** | 0 |
| 50 | 47.5 and above | 0 | 0 | 1 | 8 |

[*]Clusters are sorted based on cluster mean average speed.

**Table 4 New driving pattern classification for cycle construction**

| Class | Range of Average Travel Speed (mph) |
|-------|-------------------------------------|
| 1 | 0 ~ 12.5 |
| 2 | 12.5 ~ 32.5 |
| 3 | 32.5 ~ 47.5 |
| 4 | 47.5 and above |

As presented in Table 3, our results suggested a four-class classification scheme for arterial driving activities. Our analysis indicates that for emission modeling purposes, cycle construction should follow activity classifications that are different from those traditionally used in transportation engineering. For instance, a separate driving cycle may be desired to represent driving conditions with an average speed under 12.5 mph. In a later section of this paper, we construct individual driving cycles for each class.

*3.4    Markov-chain based cycle construction methodology*

Once cycle construction targets have been determined, we implement cycle construction in four steps: 1) snippet segmentation 2) transition matrix estimation, 3) cycle synthesis, and 4) cycle selection.

<u>*Snippet segmentation*</u>

First, we partitioned the data into modal events, or "snippets", that is, acceleration, deceleration, cruise or idle modes. These snippets are building blocks in cycle construction. To partition the continuous speed-time profiles into modal events, we used a set of parameters, $(a_1, a_2, \delta, n)$, to define the modes. An acceleration event is defined as any instantaneous point with acceleration greater than $a_1$ mph/s, or continuous observations with instantaneous accelerations greater than $a_2$ mph/s, lasting for $n$ seconds or longer, and that accumulate a speed increment greater than $\delta$ mph. Likewise, a deceleration event is defined as any instant observation with deceleration greater than $a_1$ mph/s, or continuous observations with instantaneous decelerations smaller than greater than $a_{12}$ mph/s, lasting for $n$ or more seconds, and obtaining a decrease in speed greater than $\delta$ mph. The remaining observations are classified as cruise or idle events.

To determine the optimal value of classification scheme $(a_1, a_2, \delta, n)$, we want to satisfy the following objectives. First, a modal event should comprise continuous seconds of

similar accelerations.  Therefore, $\theta_1$, the standard deviation of acceleration of any modal event should be small.  This criterion is consistent with that of Lin's MLE classifying method when applied with single clustering variable (e.g., acceleration).  In other words, the within-cluster sum of squares is part of the minimization function.  Second, for any cruise event, $\theta_2$, the difference between the maximum speed and the minimum speed, should be small.  Third, for any acceleration event or deceleration event, the absolute difference between the maximum speed and the minimum speed ($\theta_3$) should be large. Fourth, the average acceleration of any cruise event ($\theta_4$) should close to zero.  And last, because the mode of an individual point (one second of observation) highly depends on the neighboring points (seconds), we require that a good classification scheme should produce a low fraction of one-second events ($\theta_5$).

Therefore, a comprehensive assessment parameter (CAP) for the above objectives is a weighted sum,

$$CAP = k_1 \cdot \sum \theta_1 + k_2 \cdot \sum_{cruise} \theta_2 - k_3 \cdot \sum_{acc,dec} \theta_3 + k_4 \cdot \sum_{cruise} \theta_4 + k_5 \cdot \theta_5$$

The $k_1, k_2, \ldots, k_5$ are weights for each objective. In this study, all objectives are weighted equally.  We iteratively evaluated different combinations of ($a_1, a_2, \delta, n$) to obtain the smallest value of CAP.  The result showed that the minimum CAP values were obtained using a (5, 1, 3, 2) classification scheme, and therefore, we adopted these values in snippet segmentation.  In other words, we define acceleration as any instant (a single second) with acceleration higher than 5 mph/s, or any continuous observations of accelerations larger than 1 mph/s and lasting for 3 seconds or longer, and achieving a speed increment of 2 mph or more. Decelerations are defined similarly.  The (5, 1, 3, 2) classification scheme allows for representation of the natural variability in driving activity that occurs in the real-world.  In particular, the scheme allows for construction of cycles that include minor acceleration and deceleration activities during cruise operating modes; a sample trace based on this scheme is included in Figure 6.
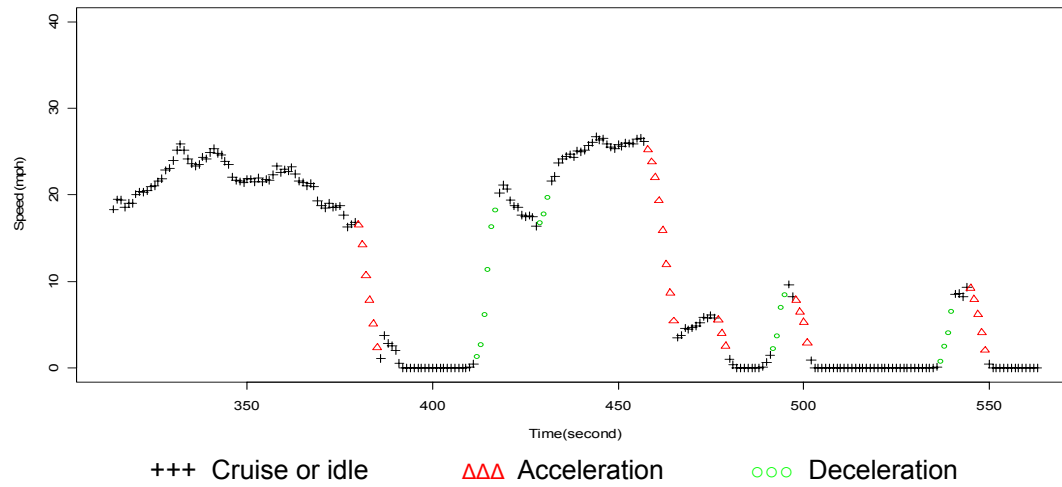
**Figure 6 Example of Snippet Segmentation Result**

Note: cruise and idle events (+ symbol) include minor acceleration and deceleration activity; sustained acceleration events are shown as green circles and sustained deceleration events are shown as red triangles.

To construct driving cycles for each cluster defined in Table 4, driving data were divided into four groups based on average speed. We define each cluster as a driving pattern "class." For each class, the snippet characteristic statistics are presented in Table 5.

**Table 5  Snippet bin characteristics by driving pattern class**

| Mode | num. of seconds | num. of snippets | mean speed | sd. Speed | mean. Acc | sd. Acc | mean VSP | sd. VSP | mean duration | sd. duration |
|---|---|---|---|---|---|---|---|---|---|---|
| Class 1 (Average speed between 0 ~ 12.5 mph) | | | | | | | | | | |
| acceleration | 7552 | 1543 | 12.79 | 9.78 | 2.70 | 1.46 | 6.75 | 5.46 | 4.89 | 2.75 |
| deceleration | 6294 | 1328 | 15.18 | 9.73 | -2.71 | 1.48 | -8.06 | 7.40 | 4.74 | 2.99 |
| cruise | 12797 | 2090 | 17.90 | 12.38 | 0.01 | 0.85 | 0.75 | 3.52 | 6.12 | 8.58 |
| idle | 31138 | 1018 | 0.16 | 0.55 | 0.01 | 0.33 | -0.01 | 0.13 | 30.59 | 27.59 |
| Class 2 (Average speed between 12.5 ~ 32.5 mph) | | | | | | | | | | |
| acceleration | 32590 | 5749 | 19.20 | 11.84 | 2.64 | 1.51 | 9.68 | 6.71 | 5.67 | 3.44 |
| deceleration | 27972 | 4901 | 22.40 | 11.99 | -2.84 | 1.56 | -11.48 | 9.80 | 5.71 | 3.72 |
| cruise | 90397 | 9552 | 32.36 | 11.38 | 0.05 | 0.76 | 2.23 | 4.92 | 9.46 | 11.86 |
| idle | 44481 | 2194 | 0.13 | 0.45 | 0.02 | 0.33 | -0.01 | 0.11 | 20.27 | 18.52 |
| Class 3 (Average speed between 32.5 ~ 47.5 mph) | | | | | | | | | | |
| acceleration | 6887 | 1313 | 29.41 | 12.27 | 2.33 | 1.36 | 13.84 | 7.59 | 5.25 | 3.37 |
| deceleration | 6136 | 1155 | 32.26 | 12.53 | -2.75 | 1.54 | -14.70 | 11.21 | 5.31 | 3.66 |
| cruise | 47698 | 2894 | 42.68 | 8.71 | 0.05 | 0.67 | 3.89 | 5.82 | 16.48 | 20.35 |
| idle | 1082 | 122 | 0.20 | 0.52 | 0.03 | 0.49 | -0.02 | 0.13 | 8.87 | 10.78 |
| Class 4 (Average speed 47.5 and above) | | | | | | | | | | |
| acceleration | 754 | 168 | 40.00 | 15.08 | 2.33 | 1.57 | 20.46 | 12.84 | 4.49 | 3.59 |
| deceleration | 733 | 166 | 44.79 | 14.15 | -2.64 | 2.04 | -17.55 | 18.55 | 4.42 | 3.24 |
| cruise | 15426 | 464 | 56.33 | 7.87 | 0.04 | 0.57 | 7.36 | 6.52 | 33.25 | 66.61 |

## Markov chain process and transition matrix

As noted earlier, our cycle construction follows the modal approach where the driving process is viewed as a Markov Chain random process (Lin and Niemeier 2002). One basic property of the Markov Chain is that the probability distribution of future states, in this case the modal nature of driving activity, depends only on the current state. In other words, the probability of a vehicle accelerating, decelerating or cruising in the next time interval depends solely on its current modal state. The probability of going from one modal state $i$ to another state $j$ are called transition probabilities; the matrix containing all transition probabilities is the transition matrix.

Let's assume a vector $\pi$, whose elements sum up to one, is the unique solution to equation $\pi = P\pi$, where $P$ is the transition matrix. Then $\pi$ is the stationary distribution of the Markov chain. For a Markov Chain with a stationary distribution, as the chain grows, the overall modal distribution will remain unchanged. Using this property, we can assume that cycles constructed by chaining modal snippets one by one using the transition matrix will result in a cycle that is representative of the overall modal distribution.

To estimate the transition matrix for each driving pattern class, we apply the maximum likelihood estimation method developed by Anderson and Goodman (1957) and restated in Lee et al. (1970). For a stationary Markov process, the probability of the ordered sequence of states can be written as follows:

$$P(X_0, X_1,..., X_T) = P(X_0)\prod_t P(X_t \mid X_{t-1})$$

If we denote $n_{ij}(t)$ as the number of individuals for which $X_t = s_j$ at time $t$ and $X_{t-1} = s_i$, the probability of a given ordered set of sequences for the $n$ observations can by written as

$$P(X_0, X_1,..., X_T \mid n) \propto P(X_0)\prod_{i,j} p_{ij}^{n_{ij}}$$

where $\propto$ denotes proportionality and $n_{ij} = \sum_t n_{ij}(t)$. By maximizing the above likelihood function with respect to the $p_{ij}$'s and subject to the constraint $\sum_j p_{ij} = 1$, the maximum likelihood estimator of transition probability is,

$$p_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$$

The estimated transition probabilities are presented in Table 6. For instance, at low speed driving conditions, following a deceleration maneuver, the vehicle is 45% likely to be engaged in cruising, and 52% likely to accelerate. For driving pattern class four, vehicles travel at high speed and generally do not idle. As a result, there are only three states in such cases.

**Table 6  Snippet Bin Transition Matrix**

| Previous Snippet Bin | Next Snippet Bin | | | |
|---|---|---|---|---|
| | Acceleration | Deceleration | Cruise | Idle |
| Class 1 (Average speed between 0 ~ 12.5 mph) | | | | |
| Acceleration | 0 | 0.06 | 0.93 | 0 |
| Deceleration | 0.03 | 0 | 0.45 | 0.52 |
| Cruise | 0.36 | 0.64 | 0 | 0 |
| Idle | 0.99 | 0.01 | 0 | 0 |
| Class 2 (Average speed between 12.5 ~ 32.5 mph) | | | | |
| Acceleration | 0 | 0.03 | 0.97 | 0 |
| Deceleration | 0.06 | 0 | 0.59 | 0.35 |
| Cruise | 0.39 | 0.61 | 0 | 0 |
| Idle | 1.00 | 0 | 0 | 0 |
| Class 3 (Average speed between 32.5 ~ 47.5 mph) | | | | |
| Acceleration | 0 | 0.01 | 0.99 | 0 |
| Deceleration | 0.06 | 0 | 0.85 | 0.09 |
| Cruise | 0.42 | 0.58 | 0 | 0 |
| Idle | 0.99 | 0.01 | 0 | 0 |
| Class 4 (Average speed 47.5 and above) | | | | |
| Acceleration | 0 | 0.04 | 0.96 | |
| Deceleration | 0.08 | 0 | 0.92 | |
| Cruise | 0.42 | 0.58 | 0 | |

*Cycle synthesis*

Our cycle synthesizing process relies on a combination of random sampling and the Markov chain approach (Figure 7). Cycle development begins with the use of snippets. A snippet is the smallest unit of activity that we track as part of the cycle-building process; a snippet can range in duration from one to 100s of seconds, and represents a single mode of operation. Once the real-world data have been partitioned into snippets representing various driving modes, the cycle development process joins snippets into a driving sequence. Each sequence is constructed as a single Markov chain (described in more detail below). The length of the driving sequence is determined by the length of the

continuous observations gathered in the field when the real-world data were obtained. For example, assume a chase vehicle tracked target vehicles and, on average, for each vehicle tracked, continuously observed two minutes of arterial driving activity; the cycle construction methodology will attempt to create driving sequences that are, on average, two minutes in duration. Sequences are then joined, as necessary, to form a complete drive cycle. A drive cycle may include one or more sequences, depending upon the desired length of the cycle. The method is implemented using the computer programming language "R" to generate candidate cycles in batches.
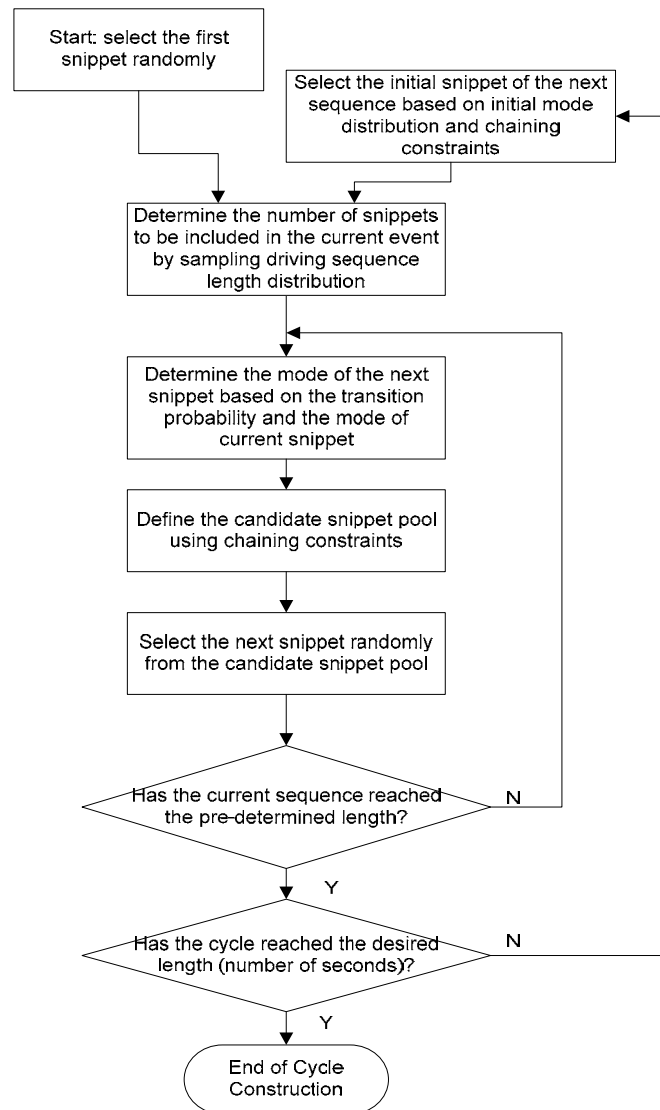


**Figure 7 Cycle Construction Program Flowchart**

18

To start the cycle development process, an initial snippet is randomly selected. Optional constraints can be imposed to require that the cycle start from zero or any speed for the convenience of dynamometer testing. Additional snippets are added one by one until the length of the current driving sequence has been met.

To add snippets, first, the mode of the next snippet is determined based on the transition probability and the current mode. Then, from the selected mode, the candidate snippet pool is defined by the chaining constraint on speed and VSP. A snippet is randomly selected from the candidate snippet pool. This process continues until the current sequence reaches the selected length. Another snippet is then randomly selected from the initial snippet pool to start a new sequence. This snippet is required to satisfy the chaining constraint given the current ending speed and VSP. The whole process repeats until the cycle reaches the desired length (desired cycle length is typically a function of dynamometer test cost considerations).

When selecting subsequent snippets from a particular bin, it is necessary that speeds and VSPs between two connecting seconds match within a certain range. We found that the 95% range of instantaneous change in speeds (i.e., instantaneous acceleration) is much wider at lower speeds and narrows as speed increases. As a result, we set the limit of instantaneous acceleration as a function of current speed as shown in Figure 8a. Specifically:

$$\text{acceleration upper bound} = \max (A + B* v, 1)$$
$$\text{deceleration lower bound} = \min (C + D* v, 1)$$

where A, B, C, D are coefficients obtained from regression, and $v$ is the ending speed of current snippet. The range of instantaneous change in VSP is relatively stable (Figure 8b). For simplicity, we therefore impose the chaining constraint that per-second change in VSP should not exceed 15 kW/tonne.
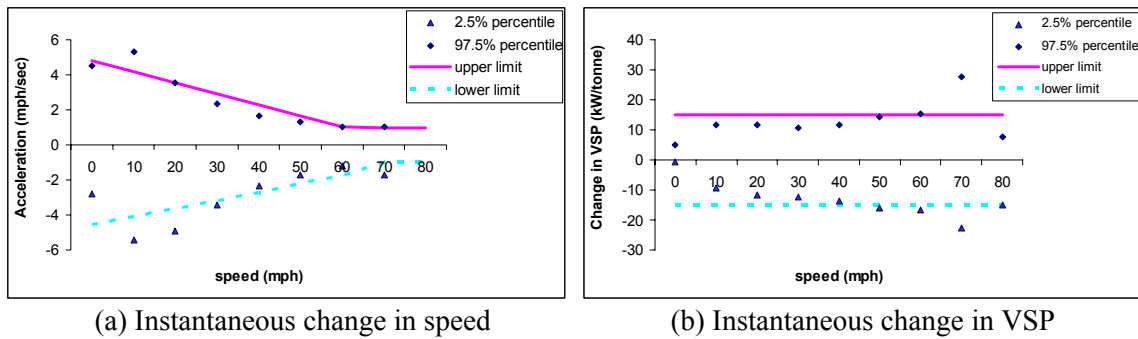


(a) Instantaneous change in speed        (b) Instantaneous change in VSP

**Figure 8 Range of Instantaneous Change in Speed and VSP**

The addition of snippets terminates when the cycle length reaches desired values. For each driving pattern (meaning the driving activity that occurred within each of the four speed bin classes shown in Table 4), more than 10,000 candidate driving cycles were constructed.

*Cycle selection*

Conventionally, cycle selection primarily relies on the sum of absolute differences between the candidate cycle SAFD and the sample SAFD (Watson et al. 1982; Gammariello and Long 1996). This particular parameter is anecdotally referred to as the "SAFDdiff". A smaller SAFDdiff suggests that a cycle is more representative of the observed sample. However, based on analyses we conducted, candidate cycles with the smallest SAFDdiff are not necessarily most representative in terms of the fraction of time spent in each mode. Therefore, we used a combination of SAFDdiff and a similarity measure of modal distribution. The SAFDdiff is defined as:

$$\text{SAFDdiff} = \sum_{a}\sum_{v} \left| fr_{(a,v)} - p_{(a,v)} \right|$$

where, $fr_{(a,v)}$ is the fraction of time spent at speed $v$ and acceleration $a$ in the candidate cycle and $p_{(a,v)}$ denotes the fraction of time spent at speed $v$ and acceleration $a$ in the observed data. The fraction of time is computed at the resolution of 1 mph for speed and 1 mph/s for acceleration in our analysis.

The similarity measure for modal distribution, or "Modediff", is calculated as the sum of absolute differences in the distribution of time across all four modes (acceleration, deceleration, idle, cruise):

$$\text{Modediff} = \sum_{m} \left| fr_m - p_m \right|$$

where $fr_m$ is the fraction of time spent in mode $m$ in the candidate cycle, and $p_m$ is the fraction of time spent in mode $m$ in the observed data.

Next, we created a statistical measure, which we call the Composite Performance Measure (CPM), to assess how well the constructed cycle matched the observed data, considering both the SAFD and the time spent in different travel modes. The CPM is the weighted sum of SAFDdiff and Mode.diff,

$$\text{CPM} = \textit{SAFDdiff} * w_1 + \textit{Modediff} * w_2$$

where the weights $w_1$ and $w_2$ are the inverse of standard deviations of SAFDdiff and Modediff, respectively.

The cycle with the smallest value of CPM was selected from each driving pattern class. We also examined the cycle with the smallest SAFDdiff measure, to facilitate comparing cycle performance based on our new technique (SAFD plus mode considerations), to cycle performance based on traditional approaches (SAFD only). Table 7 compares the candidate cycle selected based on SAFDdiff (cycle 1) and the cycle selected based on CPM (cycle2) for the low-speed driving pattern class. As expected, the cycle with the smallest SAFDdiff does not provide as good a fit of modal distribution to the observed data as does the new method. In particular, the fraction of time spent in acceleration mode is nearly 2% higher than the observed data, while the fraction of time in cruise is slightly more than 4% lower. In addition, the cycle with the smallest SAFDdiff underperforms the one with smallest CPM in the mean of (positive) acceleration and the mean of deceleration. To emphasize a key point: the importance of matching both the SAFDdiff *and* the modal distribution is demonstrated by the difference in the acceleration characteristics of the cycles (Table 7). Acceleration events are especially important to real-world emissions.

**Table 7 Comparison of Class 1 driving cycles selected based on SAFDdiff and on CPM**

|  | Observed driving | Cycle 1 (based on SAFDdiff) | | Cycle 2 (based on CPM) | |
|---|---|---|---|---|---|
| SAFDdiff |  | 0.367 |  | 0.371 |  |
| CPM |  | 33.57 |  | 32.96 |  |
| Modediff |  | 0.097 |  | 0.058 |  |
| Mean Speed (mph) | 7.38 | 6.93 | (-6.1%) | 6.76 | (-8.4%) |
| St.Dev. Speed (mph) | 10.95 | 10.58 | (-3.4%) | 10.16 | (-7.3%) |
| Mean Acceleration (mph/s) | 1.23 | 1.25 | (+1.7%) | 1.24 | (+0.6%) |
| Mean.Deceleration (mph/s) | -1.26 | -1.42 | (+12.9%) | -1.20 | (-4.1%) |
| St.Dev. Acceleration (mph/s) | 1.58 | 1.56 | (-1.1%) | 1.47 | (-6.6%) |
| % time in acceleration | 13.1 % | 15.0 % | (+1.9 %) | 0.13 | (+0.1%) |
| % time in deceleration | 10.9 % | 10.2 % | (-0.6 %) | 0.11 | (-0.1%) |
| % time in cruise | 22.1 % | 17.9 % | (-4.2 %) | 0.25 | (+2.8 %) |
| % time in idle | 53.9 % | 56.8 % | (+2.9 %) | 0.51 | (-2.8 %) |

For driving pattern class 2 and 3, the SAFDdiff measure and the CPM measure lead to the same cycle. For driving pattern class 4, we select the cycle with smallest CPM as the final cycle. A comparison between the two alternatives is presented in Table 8. The cycle selected based on the CPM measure outperforms the one selected based on SAFDdiff measure both in terms of modal distribution and speed or acceleration parameters.

**Table 8 Comparison of Class 4 driving cycles selected based on SAFDdiff and on CPM**

| | Observed driving | Cycle 1 (based on SAFDdiff) | | Cycle 2 (based on CPM) | |
|---|---|---|---|---|---|
| SAFDdiff | | 0.449 | | 0.488 | |
| CPM | | 21.85 | | 21.12 | |
| Modediff | | 0.102 | | 0.003 | |
| Mean Speed (mph) | 55.10 | 56.81 | (+3.1%) | 55.15 | (+0.1%) |
| St.Dev. Speed (mph) | 9.56 | 6.09 | (-36.3%) | 8.01 | (-16.2%) |
| Mean Acceleration (mph/s) | 0.60 | 0.43 | (-28.0%) | 0.53 | (-11.5%) |
| Mean.Deceleration (mph/s) | -0.63 | -0.45 | (-27.7%) | -0.54 | (-14.2%) |
| St.Dev. Acceleration (mph/s) | 1.06 | 0.60 | (-43.8%) | 0.86 | (-19.3%) |
| % time in acceleration | 4.5 % | 1.7 % | (-2.8%) | 4.3 % | (+0.1%) |
| % time in deceleration | 4.3 % | 2.0 % | (-2.3%) | 4.3 % | (+0.0%) |
| % time in cruise | 91.2 % | 96.3 % | (+5.1%) | 91.4 % | (+0.2%) |

Note that there is no idle mode in the speeds represented by the Class 4 bin.

## 4. Results

### 4.1 Final cycles

A final cycle was selected from over 10,000 candidate cycles for each driving pattern class. Speed-time traces of four final cycles are presented in Figures 9 through 12.
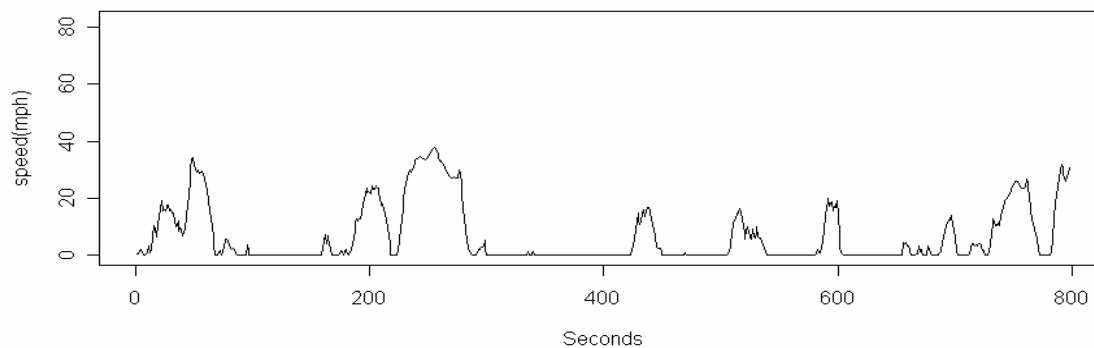


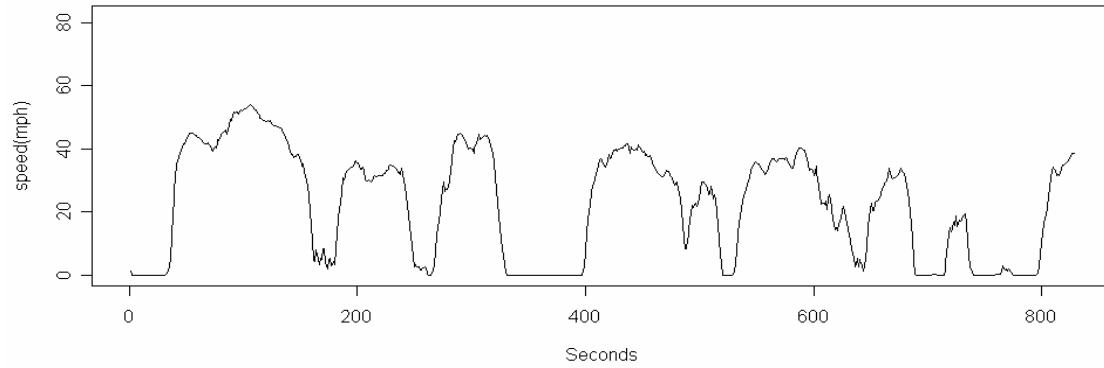**Figure 9 Final Arterial Driving Cycle for Ave. Speed 0 ~ 12.5 mph (driving pattern Class 1)**

**Figure 10 Final Arterial Driving Cycle for Ave. Speed 12.5 ~ 32.5 mph (Class 2)**
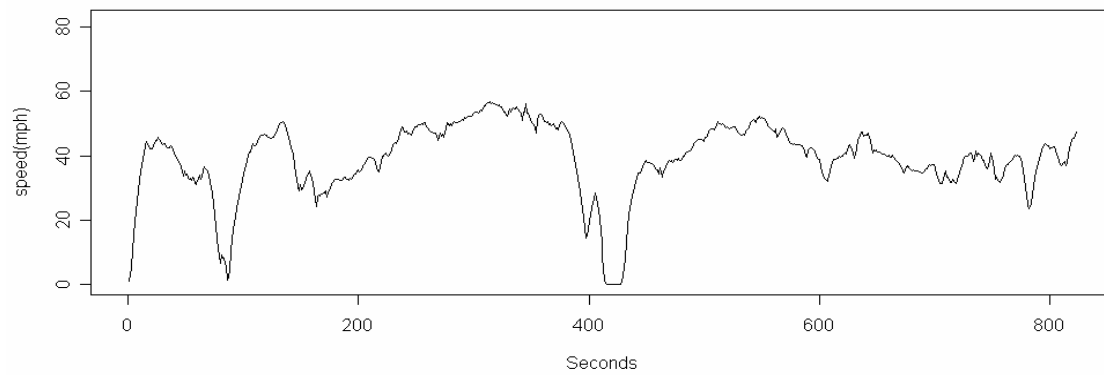


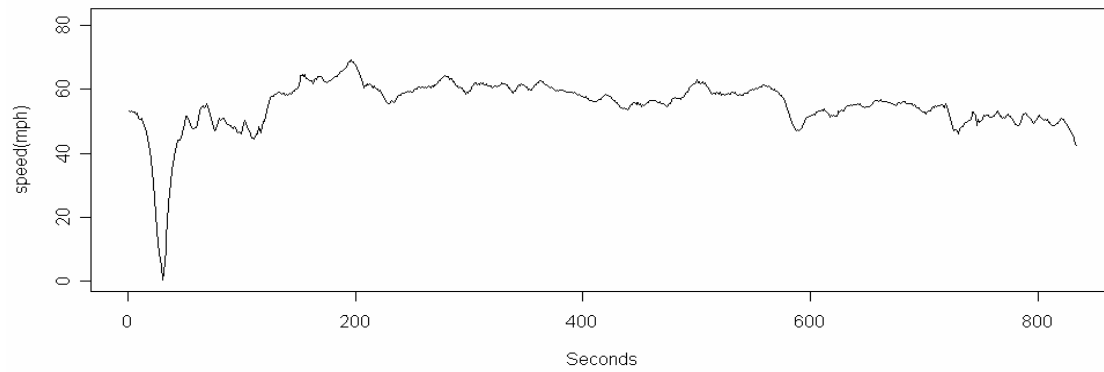**Figure 11 Final Arterial Driving Cycle for Ave. Speed 32.5 ~ 47.5 mph (Class 3)**



**Figure 12 Final Arterial Driving Cycle for Ave. Speed 12.5 ~ 37.5mph (Class 4)**

23

As shown in Figure 9, the lowest speed driving cycle (between 0 to 12.5 mph) contains more idling, stop-and-go incidents, and brief acceleration-deceleration maneuvers. These maneuvers are most likely associated with congestion or traffic impedance (e.g., traffic lights). The driving cycle for speeds between 12.5 to 32.5 mph in Figure 10 also contains a significant fraction of idling but exhibits a higher cruising speed. The medium speed driving cycle (Figure 11) has only one full stop and two substantial deceleration-acceleration maneuvers. The high speed arterial driving cycle in Figure 12 is mostly smooth flow occurring at or above 45 mph, with the exception of a single interruption. We performed a chi-square ($\chi^2$) test on the modal distributions of activity for each cycle compared to the observed driving data. The results supported that cycle modal distribution is consistent with the observed driving data. The cycle length, SAFDdiff and chi-square test statistics on the modal distribution are presented in Table 9.

**Table 9 Final Driving Cycle Length, SAFDdiff, and $\chi^2$ test statistics**

| Driving Pattern Class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Cycle Duration (sec) | 798 | 829 | 823 | 834 |
| Cycle Length (mile) | 1.50 | 5.18 | 8.92 | 12.78 |
| SAFDdiff | 0.371 | 0.553 | 0.515 | 0.488 |
| CPM | 32.96 | 24.63 | 23.61 | 21.12 |
| $\chi^2$ test | | | | |
| $\chi^2$ statistics | 3.9487 | 3.5327 | 1.3469 | 0.0404 |
| p-value | 0.27 | 0.32 | 0.72 | 0.98 |
| df | 3 | 3 | 3 | 2 |

*4.2    Comparison between observed driving characteristics and driving cycles*

We also compared a number of different parameters between the observed driving data and the final cycles: average speed, standard deviation of speeds, average positive acceleration, average negative acceleration, standard deviation of acceleration, average positive VSP, average negative VSP, standard deviation of VSP, positive kinetic energy (PKE), and percentages of time (%) spent in each mode, respectively for acceleration, deceleration, cruise, and idle. The comparison is presented in Table 10.

In general, our cycles replicate the modal distribution well. Some statistics, especially the mean of negative acceleration and VSP and the standard deviation of the acceleration and VSP are smaller than the observed data. This is likely due to the heavier tails observed on both ends of the driving activity data. Figures 13 to 16 illustrate the SAFD of the original target vehicle data, the SAFD of the final cycles and their difference in 3D plots for the four driving pattern classes.

**Table 10 Comparison of Cycle performance against observed driving data**

| | Class 1 | | | Class 2 | | | Class 3 | | | Class 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Obs. | Cycle | % Diff | Obs. | Cycle | % Diff | Obs. | Cycle | % Diff | Obs | Cycle | % Diff |
| Mean speed (mph) | 7.4 | 6.8 | -8% | 21.4 | 22.5 | 5% | 39.4 | 39.0 | -1% | 55.1 | 55.2 | 0% |
| Std. speed (mph) | 11.0 | 10.2 | -7% | 16.2 | 17.2 | 6% | 11.9 | 11.3 | -6% | 9.6 | 8.0 | -16% |
| Mean acceleration (mph/s) | 1.2 | 1.2 | 1% | 1.3 | 1.3 | -2% | 0.9 | 0.9 | 2% | 0.6 | 0.5 | -12% |
| Mean.deceleration (mph/s) | -1.3 | -1.2 | -4% | -1.4 | -1.2 | -13% | -1.0 | -0.8 | -15% | -0.6 | -0.5 | -14% |
| Std. acceleration (mph/s) | 1.6 | 1.5 | -7% | 1.8 | 1.6 | -10% | 1.5 | 1.2 | -15% | 1.1 | 0.9 | -19% |
| Mean VSP(+) | 3.0 | 2.9 | -5% | 5.7 | 5.6 | -1% | 7.3 | 7.1 | -3% | 9.3 | 8.6 | -8% |
| Mean VSP(-) | -4.4 | -3.6 | -19% | -6.7 | -5.8 | -14% | -7.3 | -5.8 | -20% | -8.1 | -5.0 | -38% |
| Std. VSP | 5.1 | 4.2 | -17% | 8.3 | 7.1 | -15% | 9.5 | 7.9 | -17% | 9.8 | 7.2 | -26% |
| PKE (mile/hr$^2$) | 1.3 | 1.6 | 24% | 1.0 | 1.1 | 6% | 0.7 | 0.8 | 11% | 0.5 | 0.5 | -11% |
| Median acceleration | 0.7 | 0.8 | 21% | 0.8 | 0.8 | 1% | 0.6 | 0.7 | 12% | 0.4 | 0.3 | -12% |
| Median decelaration | -0.8 | -0.9 | 16% | -0.8 | -0.7 | -14% | -0.6 | -0.5 | 0% | -0.4 | -0.4 | -3% |
| Median VSP(+) | 0.8 | 1.0 | 34% | 4.4 | 4.7 | 5% | 6.2 | 6.3 | 2% | 8.5 | 7.8 | -8% |
| Median VSP(-) | -2.1 | -2.0 | -1% | -4.0 | -4.0 | -2% | -4.3 | -4.4 | 3% | -4.4 | -3.3 | -24% |
| | | | | | | | | | | | | |
| Distribution of time | | | Diff | | | Diff | | | Diff | | | Diff |
| Acceleration (%) | 13.1 | 13.2 | 0.1 | 16.7 | 15.0 | -1.7 | 11.1 | 12.4 | 1.3 | 4.5 | 4.3 | -0.1 |
| Deceleration (%) | 10.9 | 10.8 | -0.1 | 14.3 | 13.0 | -1.3 | 9.9 | 9.7 | -0.2 | 4.3 | 4.3 | 0.0 |
| Cruise (%) | 22.1 | 24.9 | 2.8 | 46.3 | 48.5 | 2.2 | 77.2 | 76.1 | -1.1 | 91.2 | 91.4 | 0.2 |
| Idle (%) | 53.9 | 51.1 | -2.8 | 22.8 | 23.5 | 0.8 | 1.8 | 1.8 | 0.1 | 0.0 | 0.0 | 0.0 |

**Figure 13 Cycle SAFD and Observed SAFD for Average Speed between 0 ~ 12.5 mph**

a) Target (observed)
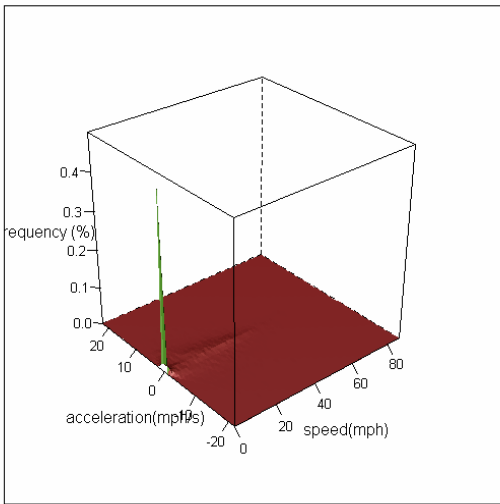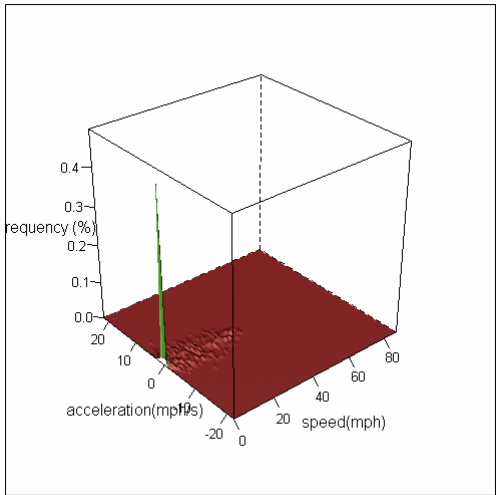
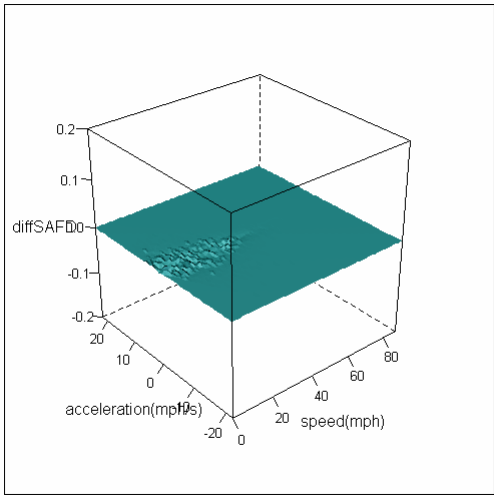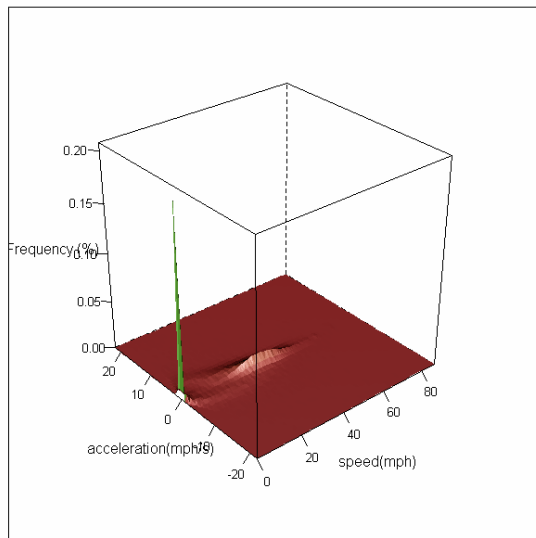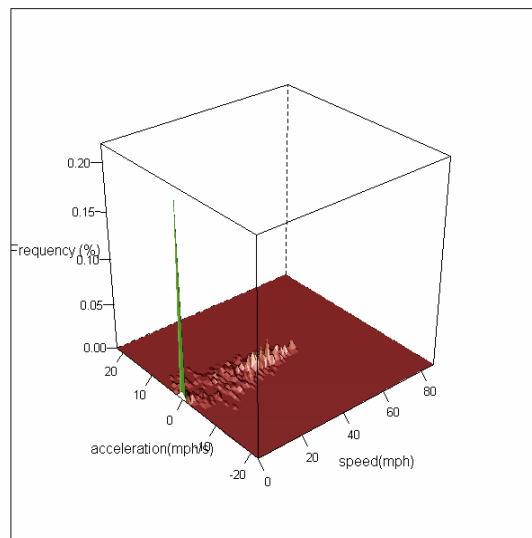b) Final cycle

c) Difference

**Figure 14 Cycle SAFD and Observed SAFD for Average Speed between 12.5 ~ 32.5 mph**

a) Target (observed)

b) Final cycle

c) Difference

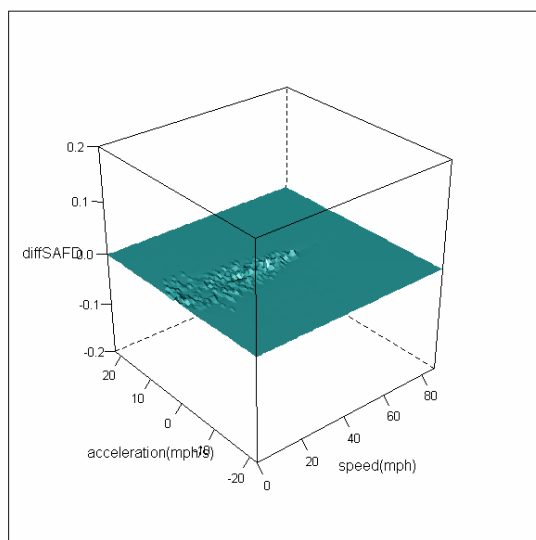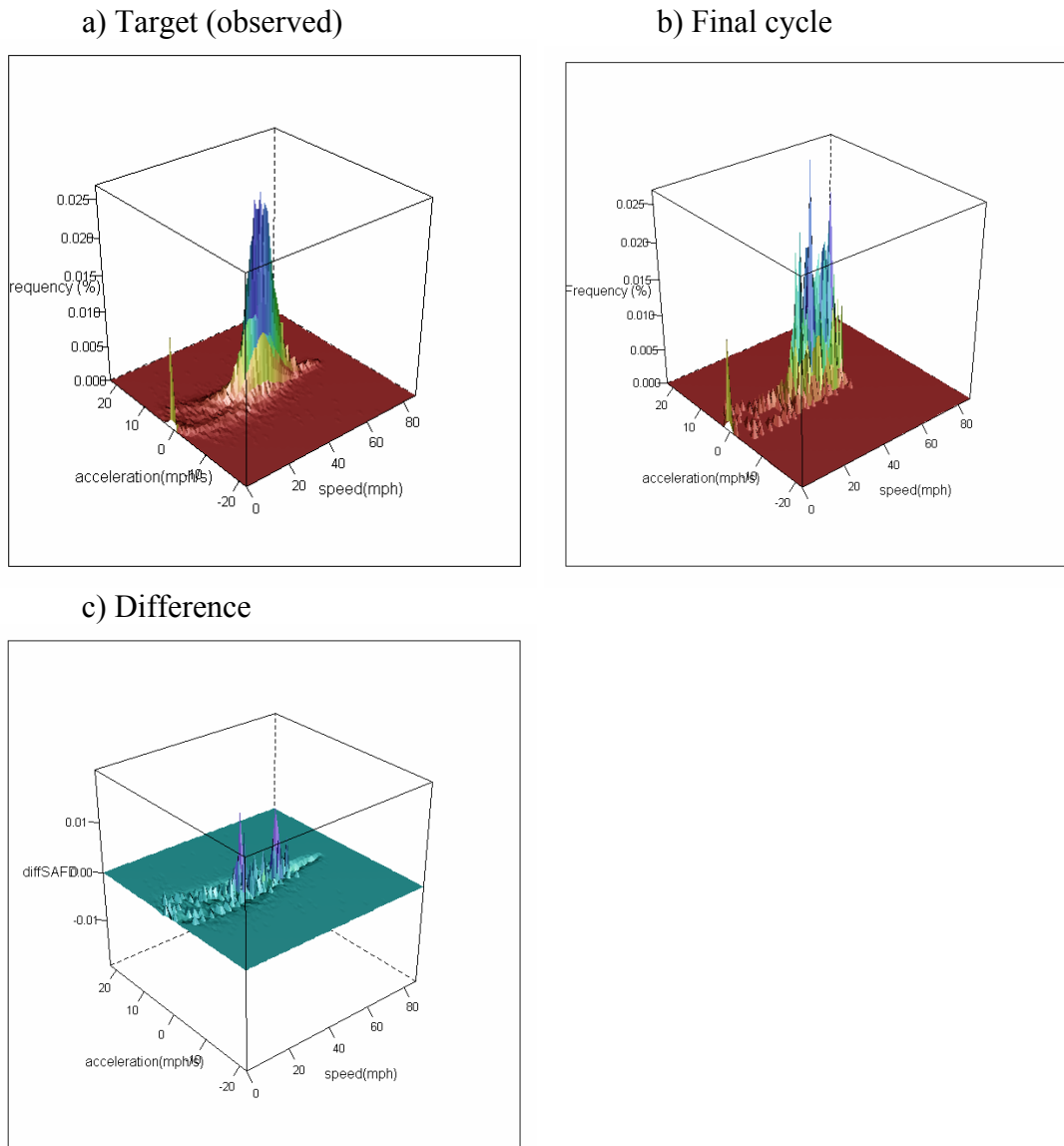**Figure 15 Cycle SAFD and Observed SAFD for Average Speed between 32.5 ~ 47.5 mph**
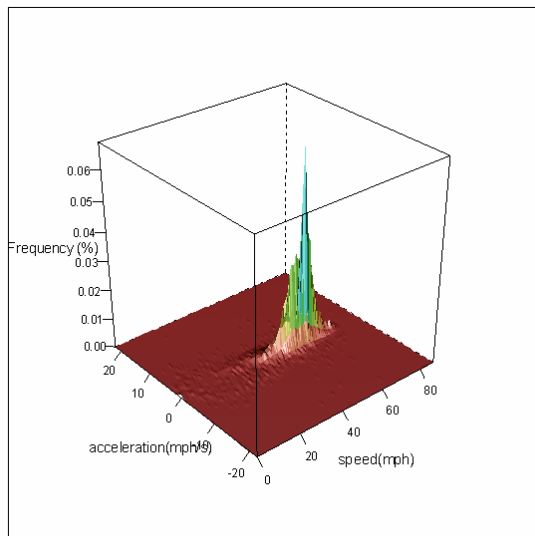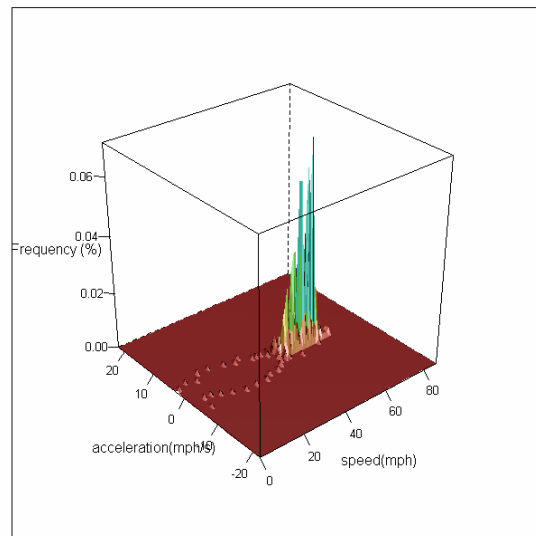
a) Target (observed)



b) Final cycle



c) Difference

**Figure 16 Cycle SAFD and Observed SAFD for Average Speed 47.5 mph and above**
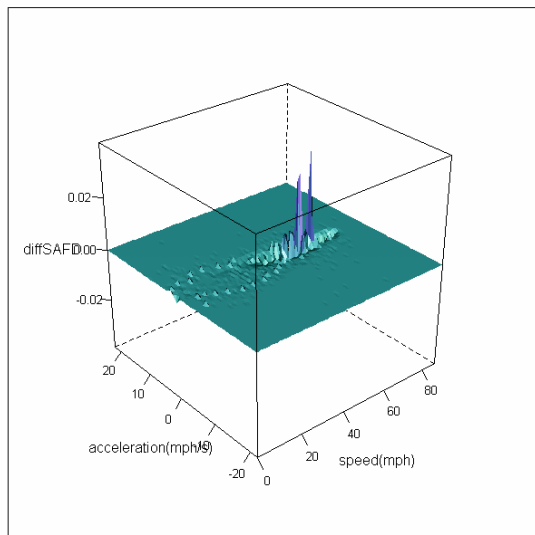
a) Target (observed)

b) Final cycle





c) Difference

## 5. Conclusions

This study presented a systematic driving cycle development method which includes two components. First, it includes a statistical analysis of observed driving patterns to identify cycle construction targets. Instead of using existing traffic classifications that were not designated for the purposes of emission estimation, this study suggests that target driving patterns should first be defined using emission-related variables. In this study, we classified driving patterns using average speed and VSP bin distribution. Our study showed that a statistical classification result differs from existing approaches that rely on LOS classification (i.e., approaches that focus on matching SAFDdiff only, rather than also considering modal distributions). In particular, independent arterial cycles should be constructed to represent driving patterns at very low average speeds (below 12.5 mph), at moderate average speeds (12.5-32.5 mph, and 32.5-47.5 mph) and at very high average speeds (above 47.5 mph). Classification results were used to direct arterial cycle construction.

This study also includes a modal approach in cycle development. This method is developed from the Markov chain cycle construction method first introduced by Lin and Niemeier (2002). First, driving traces were segmented into small modal snippets using specific criteria to define driving mode. Then cycles were constructed as a sequence of modal snippets. The candidate cycles were assessed by comparing observed and constructed data using modal measures in addition to traditional SAFD. The resulting cycles replicated the modal distribution very well.

In addition to introducing a new cycle construction approach, this work also introduces a new test statistic to assess cycle performance: the Composite Performance Measure, or CPM. The CPM measures how well a cycle matches the SAFD and modal distributions of real-world data. It can be used to facilitate comparisons of cycles created using the methods presented here, as well as comparisons of cycles created using different cycle development methods.

Cycles representing the two mid-range speed activity clusters performed equally well regardless of whether they were created with the new techniques presented here or with a traditional approach. However, cycles created for the lowest and highest speed activity clusters performed better at representing real-world data when they were constructed to match SAFD *and* mode distributions, rather than SAFD metrics alone. As shown in Tables 7 and 8, the percentage difference between observed and cycle-constructed mean accelerations was a factor of two to three smaller for the low- and high-speed cycles created using SAFD and mode distribution metrics. Given the relationship between

accelerations and emissions, the new technique improves representation of real-world activity that affects emissions.

Our method constructs cycles using a combination of random sampling strategies. It is designed to replicate the distribution of modal snippets and the transition probability between modes. As a result, it is designed to optimize representation of activity associated with emissions, rather than to optimize observed-to-estimated driving activity in general. The cycle construction process requires a large amount of computing resources. In addition, our mode-based perspective evaluates representativeness differently than conventional approaches. Greater weight is given to modal distribution rather than parameters such as average speed or average acceleration.

Future research could investigate whether alternative evaluation parameters improve prediction of activity that affects real-world emissions. In particular, the approach presented here could be refined in the future by completing additional analyses using the same methodology.  Further analyses could employ additional vehicle-specific data, or use of posted speed limits or more accurate measurements of traffic speed to represent average speeds.

# 6. References

An, F., Barth, M., and Ross, M. H. (1998). "Modeling enleanment emissions for light-duty vehicles." *Transportation Research Record*, 1641, 48-57.

Anderson, T. W., and Goodman, L. A. (1957). "Statistical Inference about Markov Chain." *The Annals of Mathematical Statistics*, 28(1), 89-110.

Andre, M. (1996). "Driving cycle development: characterization of the methods." SAE Technical Paper Series, the Engineering Society for Advanced Mobility, Detroit, Michigan.

Andre, M. (2004). "The ARTEMIS European driving cycles for measuring car pollutant emissions." *Science of The Total Environment*, 334-335, 73.

Andre, M., Hickman, J., Hassel, D., and Joumard, R. (1995). "Driving cycles for emission measurements under european conditions." SAE Technical Paper Series, the Engineering Society for Advanced Mobility, Detroit, Michigan.

Andre, M., and Rapone, M. (2008). "Analysis and modelling of the pollutant emissions from European cars as regards the driving characteristics and test cycles." *Atmospheric Environment*, In Press, Accepted Manuscript.

Austin, T. C., Carlson, T. R., and Dulla, R. G. (1995). "Methodology for Generating Driving Cycles for Inventory Development."

Austin, T. C., DiGenova, F. J., Carlson, T. R., Joy, R. W., Gianolini, K. A., and Lee, J. M. (1993). "Characterization of driving patterns and emissions from light-duty vehicles in California." The final report, Contract No. A932-185, California Air Resources Board, Sacramento.

Barth, M., An, F., and Norbeck, J. (1996). "Modal emissions modeling: a physical approach." *Transportation Research Record*, 1520, 81-88.

Carlson, T. R., and Austin, T. C. (1997). "Development of Speed Correction Cycles." *SR97-04-01*, Sierra Research, Inc., Sacramento, California.

Eisinger, D. S., Niemeier, D. A., Stoechenius, T., Kear, T. P., Brady, M. J., Pollack, A. k., and Long, J. (2006). "Collecting Driving Data to Support Mobile Source Emissions Estimation." *Journal of Transportation Engineering*, 132(11), 845-854.

EPA (2004). "MOVES2004 Software Design Reference Manual Draft." *EPA420-P-04-017*, Office of Transportation and Air Quality, U.S. Environmental Protection Agency.

EPA (2005). "MOVES2004 Energy and Emission Inputs Draft Report." *EPA420-P-05-003*, Office of Transportation and Air Quality, U.S. Environmental Protection Agency.

Fomunung, I., Washington, S., and Guensler, R. (1999). "A statistical model for estimating oxides of nitrogen emissions from light duty motor vehicles." *Transportation Research Part D: Transport and Environment*, 4(5), 333-352.

Gammariello, R. T., and Long, J. R. (1996). "Development Of Unified Correction Cycles." *Sixth CRC On-Road Vehicle Emissions Workshop*, San Diego, California.

Goodwin, R. W., and Ross, M. H. (1996). "Off-Cycle Exhaust Emissions from Modern Passenger Cars with Properly-Functioning Emission Controls." *SAE Technical Paper Series*, 960064.

Hung, W.-T., Tam, K.-M., Lee, C.-P., Chan, L.-Y., and Cheung, C.-S. (2005). "Comparison of driving characteristics in cities of Pearl River Delta, China." *Atmospheric Environment*, 39(4), 615.

Kent, J. H., Allen, G. H., and Rule, G. (1978). "A driving cycle for Sydney." *Transportation Research*, 12(3), 147.

Lee, T. C., Judge, G. G., and Zellner, A. (1970). *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*, North-Holland Publishing Company, Amsterdam, London.

Lin, J., and Niemeier, D. A. (2002). "An exploratory analysis comparing a stochastic driving cycle to California's regulatory cycle." *Atmospheric Environment*, 36(38), 5759.

Lin, J., and Niemeier, D. A. (2003). "Estimating regional air quality vehicle emission inventories: constructing robust driving cycles." *Transportation Science*, 37(3), 330-346.

Morey, J. E., Limanond, T., and Niemeier, D. A. (2000). "Validity of Chase Car Data Used in Developing Emission Cycles." *Journal of Transportation and Statistics*, 3(2).

SPSS (2001). "The SPSS TwoStep Cluster Component -A scalable component enabling more efficient customer segmentation."

Watson, H. C., Milkins, E. E., and Braunsteins, J. (1982). "The Development of the Melbourne Peak Cycle." *SAE/ARRB 2nd Conference on Traffic Energy and Emissions*, Melbourne.